

FILIÈRES AGRICOLES & ALIMENTAIRES

Enjeux et problématique de la numérisation



Numérisation de la recherche

Sommaire

Chapitre V : Numérisation de la recherche

1 - CONTEXTUALISATION

1.1 - Une numérisation de la société qui questionne

2 - NUMÉRISATION DE LA RECHERCHE & COLLECTE DES DONNÉES

2.1 - Intégrer le numérique dans les protocoles et les collaborations de recherche

2.2 - L'exemple des recherches menées sur la base des données disponible avec Open Food Facts

La soutenabilité du système alimentaire questionnée par les produits ultra-transformés

2.3 - La numérisation des protocoles : Nutrinet-Santé

2.4 - L'hybridation des méthodes classiques et numériques dans les protocoles : le CREDOC

Le baromètre CAF: Comportements et attitudes alimentaires en France

2.5 - Réflexions sur les conséquences

Le risque de la fracture numérique

La perte d'une compétence de la recherche : formation et accompagnement des participants

3 - LA NUMÉRISATION AU SERVICE DE LA RECHERCHE PARTICIPATIVE

3.1 - Capitaliser sur les possibilités du numérique pour faire avancer la recherche participative sur la biodiversité

3.2 - Démocratisation des outils numériques : The Wisconsin Bat Program

3.3 - Education des participations à la biodiversité : Vigie-Nature et l'OAB

3.4 - Le traitement de la donnée par les participants : The Penguin Watch Project

3.5 - Réflexions sur les conséquences

La qualité de la donnée

Portée pédagogique et éducative de la participation

Le traitement de la donnée : l'humain, le machine learning et l'intelligence artificielle

4 - NUMÉRISATION DE LA RECHERCHE : ABSENCE ET SURABONDANCE DE DONNÉES

4.1 - Dépendance de la recherche à l'existence des données et à leur traitement par les outils numériques

4.2 - Pallier le manque de données : les collectifs citoyens sur les produits phytosanitaires

4.3 - Aider à l'interprétation des données : les Adverse Outcome Pathways

4.4 - Réflexions sur les connaissances

BIBLIOGRAPHIE

1 - CONTEXTUALISATION

1.1 - Une numérisation de la société qui questionne

La numérisation de la société impacte la quantité de données produites et disponibles pour être collectées :

- **Le numérique est devenu omniprésent** dans notre vie quotidienne et est au cœur d'une multitude de petites actions individuelles : recherche d'information en ligne, consultation de sites internet, démarches administratives, espace de travail partagé, calcul d'itinéraire etc.
 - Toutes ces actions sont génératrices de données numériques dont le potentiel n'est pas nécessairement exploité mais peut l'être : autrement dit, **ces données numériques peuvent intéresser différents acteurs qui vont les valoriser afin de renforcer et/ou de développer leurs activités.**
- **Corollaire de ce constat, les objets numériques sont largement diffusés et démocratisés.** Ils peuvent être intégrés à notre quotidien afin de notamment nous aider à remplir les actions listées ci-dessous, ou être dédiés et donc générés un flux spécifique de données numériques

Cette numérisation de la société interroge la recherche (publique, privée et citoyenne) sur deux grands aspects qui la caractérisent : **sa capacité d'observation et sa capacité d'expérimentation** (1). En effet :

- **De nouvelles données à observer par la recherche sont générées.** Il s'agit soit de données plus nombreuses ou de meilleure qualité sur des sujets déjà observés précédemment, soit de nouvelles données qui n'existaient pas avant.
- **De nouveaux procédés de traitement des données se développent** et ouvrent de nouvelles possibilités **d'expérimentation** pour la recherche.

Dans ce contexte, trois grands enjeux se posent à la recherche et l'obligent à se transformer (1) :

1 - Expansion des champs d'étude et des capacités de la recherche. Comme dit précédemment, la numérisation de la société permet à la fois de **générer de nouvelles et/ou plus nombreuses données à observer et investiguer pour la recherche**, mais aussi de **renforcer voire développer de nouvelles capacités d'analyse grâce aux compétences numériques** (outils, algorithmes etc.). La surface d'investigation et des connaissances de la recherche est étendue.

2 - Restructuration des programmes de recherche. La numérisation facilite **les échanges et les utilisations croisées des données**. En cela, elle incite et favorise les travaux de **recherche associant différents acteurs** : soit plusieurs disciplines de recherche académique, soit organisations de recherche académique avec des organisations de recherche citoyenne voire des organisations privées ou associatives.

3 - En écho aux enjeux sociétaux actuels. En conséquence des deux enjeux précédents, la recherche peut investiguer et analyser de **nouveaux sujets potentiellement d'intérêt fort pour la société et s'associer plus étroitement avec des acteurs non-chercheurs** – ce tant dans la collecte que dans la diffusion de ses résultats.

- ➔ Dans la suite de cette présentation, nous allons préciser et entrer dans le détail de ces trois enjeux :
 - Ces trois enjeux se recoupent dans les différents exemples à suivre.
 - Ces exemples ont été sélectionnés afin d'illustrer les impacts concrets de la numérisation sur la recherche et en analyser les conséquences.

(1) Réseau PROSPER, « Big & Open data en recherche à l'horizon 2040 », 2019

2 - NUMÉRISATION DE LA RECHERCHE & COLLECTE DES DONNÉES

2.1 - Intégrer le numérique dans les protocoles et les collaborations de recherche

Nous allons tout d'abord discuter ici de l'intégration de la dimension numérique dans les programmes de recherche à deux niveaux :

- **Le numérique dans les protocoles de recherche.** Les programmes de recherche peuvent compléter les méthodes classiques (par exemple des cahiers d'observation papier tenus par les participants et envoyés ensuite par mail) mises en place dans leur protocole par un recours aux technologies (par exemple la saisie informatique) – voire à l'extrême, que ces dernières remplacent intégralement les méthodes classiques.
 - **La collaboration de la recherche avec des acteurs tiers du numérique.** Les nouvelles capacités d'observation et de collecte de données générées par le numérique ne sont pas exclusives à la recherche académique. Aujourd'hui, de nouveaux acteurs peuvent également utiliser ces données pour développer leur propre activité économique. Ils peuvent être des concurrents de la recherche sur son « pré-carré » ou à l'inverse des collaborateurs (1).
- Nous allons approfondir ces deux aspects ci-après avec les exemples des études menées sur la base d'Open Food Facts, du programme de recherche Nutrinet, et enfin du Centre de recherche pour l'étude et l'observation des conditions de vie (CREDOC).

- Les pages suivantes présentes tout d'abord les exemples retenus avant d'analyser et discuter leurs conséquences.

(1) Entretien avec une chargée d'études du CREDOC

(2) Open Food Facts est étudié ici uniquement sous l'angle de l'utilisation de ses données par la recherche. Le projet de Open Food Facts d'informer tout à chacun sur la composition des produits dans un objectif d'éducation sur les questions de nutrition, alimentation et santé est abordé dans la présentation PPT dédiée à la consommation de cette même étude. Pour en savoir plus sur Open Food Facts :

EN SAVOIR PLUS

(3) Exemple sur un produit :

EN SAVOIR PLUS

(4) Selon les résultats de recherche disponibles sur Google Scholar en mars 2021

(5) A. Ngeuilbaye et al., « Adoption of human metabolic processes as Data Quality Based Models », 2020 ; A. Ramirez-Portilla et al., « The era of Open Food? Exploring the influence of open and collaborative innovation in the food industry », 2016

(6) B. Soutjis, « Gouverner la qualité alimentaire par les applications », 2020 ; F. Cochoy et al., « Digitalizing consumer society: equipment and devices of digital consumption », 2020

(7) P. Loupiac et C. Blutz, « Au croisement du marketing et de la comptabilité : l'essor des activités d'évaluation et de contrôle par le consommateur », 2020

2.2 - L'exemple des recherches menées sur la base des données disponible avec Open Food Facts

Le premier exemple est celui des **recherches menées à partir des données numériques collectées par Open Food Facts**, illustrant les possibilités nouvelles de champ de recherche permises par la numérisation de la société.

Open Food Facts est une base de données sur les produits alimentaires commercialisés dans le monde collaborative, libre et ouverte (2).

Pour chaque produit référencé (à date, environ 700 000), **Open Food Facts renseigne de nombreuses caractéristiques « produit »** (allant du conditionnement au lieux de fabrication ou de vente (3)), comprenant les informations nutritionnelles et sur les additifs alimentaires.

Mise à disposition de tout à chacun, cette base de données a été depuis 2012 à l'origine de nombreuses études.

A date, plus de 200 articles (4) scientifiques publiés ont notamment utilisé Open Food Facts sous différentes approches disciplinaires : santé/nutrition (voir exemple ci-contre) mais aussi sciences et technologies (5), sociologie (6) et économie management (7).

La soutenabilité du système alimentaire questionnée par les produits ultra-transformés

Récemment, les données observées et collectées via la base de données Open Food Facts ont été utilisées dans le cadre de recherches menées par l'INRAE sur les impacts des produits ultra-transformés (1).

Plus précisément, les impacts de la consommation de ces produits ultra-transformés étaient analysés sur l'environnement, le bien-être animal, la biodiversité et d'autres dimensions culturelles et socio-économiques.

Le traitement de la base de données a permis pour chaque ingrédient additif et non-additif utilisé dans l'ultra-transformation de quantifier le nombre de produits.

Une fois les ingrédients identifiés, les chercheurs ont pu recomposer et associer les modes de production agricole majoritairement utilisés pour en conclure que les produits ultra-transformés sont principalement associés à de l'élevage et de l'agriculture intensifs.

2.3 - La numérisation des protocoles : Nutrinet-Santé

Le deuxième exemple choisi est celui de Nutrinet-Santé qui nous permet de comprendre comment le numérique s'intègre désormais au protocole de recherche.

Nutrinet-Santé est une étude épidémiologique sur les comportements alimentaires et les relations entre nutrition et santé (2). Plus spécifiquement elle étudie :

- Les **comportements alimentaires et leurs déterminants** en fonction de l'âge, du sexe, des conditions socio-économiques, du lieu de vie etc.
- Les **relations entre les apports alimentaires, l'activité physique, l'état nutritionnel et la santé** afin d'informer les grands problèmes de santé : obésité, hypertension artérielle, diabète, cancers, vieillissement etc.

C'est une étude de cohorte prospective d'observation (3) :

- Lancée en 2009, renouvelée pour 10 ans en 2013
- 500.000 participants (« nutrinautes ») de plus de 18 ans Site internet dédié regroupant les informations sur l'étude et la saisie des informations via les questionnaires numérisés

Elle a une forte dimension d'utilité publique (4) :

- Comprendre les **facteurs de risque ou de protection** pour les maladies
- Établir des **recommandations nutritionnelles** permettant de prévenir le risque et améliorer la qualité de la santé (population actuelle et future)

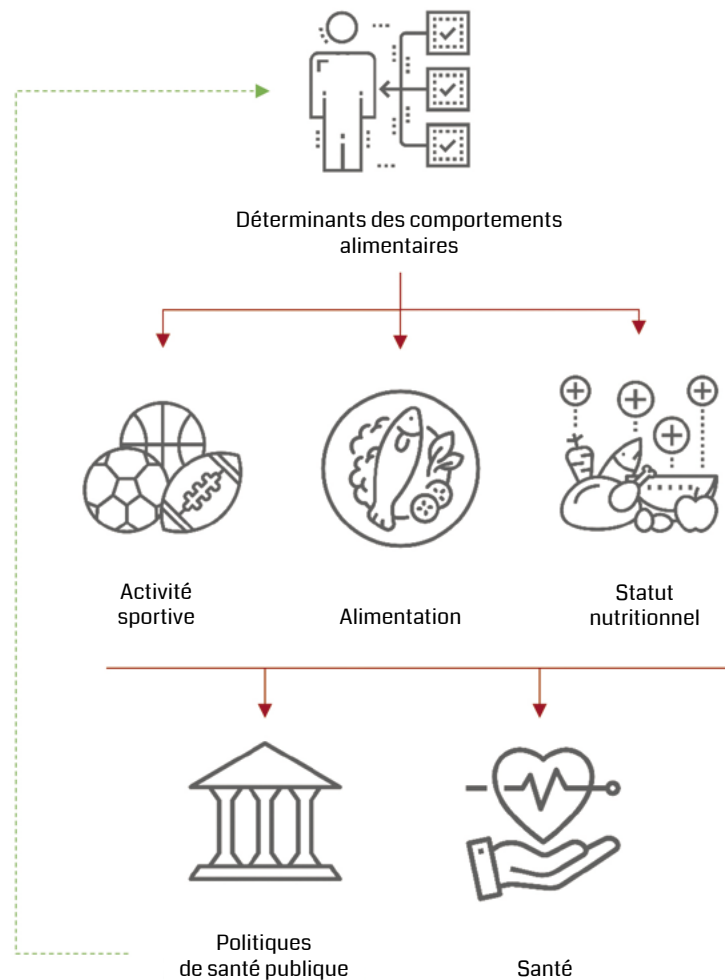
(1) A. Fardet et E. Rock, « Ultra-processed foods and food system sustainability: what are the links ? », 2020

(2) Étude Nutrinet-santé

EN SAVOIR PLUS

(3) CNAM, « Les nouveaux résultats de l'enquête Nutrinet-Santé décryptent plus précisément les comportements alimentaires des français », 2011

(4) Chantal J. et al., « Contribution of ultra-processed foods in the diet of adults from the French NutriNet-Santé study », 2017



Schématisation
du fonctionnement
de l'étude
Nutrinet-Santé,
BASIC d'après
Nutrinet-Santé

2.4 - L'hybridation des méthodes classiques et numériques dans les protocoles : le CREDOC

Le troisième exemple retenu est celui du CREDOC car il vient questionner **l'hybridation entre méthodes classiques et numériques** dans les protocoles de recherche ainsi que les **opportunités de collaboration avec des acteurs tiers**.

Le CREDOC (Centre de recherche pour l'étude et l'observation des conditions de vie) est un **organisme d'études et de recherches sur les comportements individuelles**.

Au-delà des études ponctuelles pour des demandes et clients précis, le CREDOC a depuis sa création un dispositif permanent d'enquête sur le modes de vie et opinions des français (1).

Le CREDOC est le seul institut en France à avoir de la donnée sur le temps long au sujet des comportements alimentaires des français (2).

- L'ANSES ayant changé de méthodologie il y a quelques années, **seul le CREDOC avec son baromètre développé en 1988 peut renseigner les évolutions des comportements alimentaires** des français sur les plus de 30 dernières années (3) et voir exemple ci-contre).

- Les **enquêtes** de ce baromètre sont menées auprès d'un échantillon représentatif de 3000 personnes de plus de 15 ans via notamment des carnets alimentaires hebdomadaires (**papier et également numériques depuis 2013**).

(1) credoc.fr

EN SAVOIR PLUS

(2) Entretien avec une chargée d'études du CREDOC

(1) Ibid ; CREDOC,
« Comportements et attitudes
alimentaires en France - CAF
2021 », 2021

Le baromètre CAF : Comportements et attitudes alimentaires en France

Créé en 1988, le baromètre CAF suit les comportements de consommation alimentaire des français en étudiant :

- d'un côté, leurs modes d'approvisionnement, les critères d'achat ou encore l'imaginaire lié à la nourriture
- de l'autre, le contexte de l'alimentation : rythme, pratique de la cuisine etc.

Depuis 2021, **une collaboration a été mise en place entre l'application myLabel** et ce baromètre afin de rajouter une couche de contextualisation et d'analyse supplémentaire des données observées et collectées via cette enquête.

Cette collaboration est à double-sens (1) :

- le CREDOC vient ainsi renforcer la capacité d'analyse du baromètre en incluant les données de myLabel sur les valeurs associées aux choix alimentaires
- myLabel bénéficie des résultats du baromètre

2.5 - Réflexions sur les conséquences

Le risque de la fracture numérique

Les exemples de Nutrinet-Santé et du baromètre CAF conduit par le CREDOC illustrent un questionnement de la recherche sur les protocoles et les méthodes de collecte de données : entre **totale numérisation et hybridation entre techniques classiques et numériques dans les protocoles**

- L'étude Nutrinet (2) a fait le choix d'un **total recours à la numérisation**, permettant d'entrer dans un grand **niveau de détails élevé et harmonisé** des informations saisies dans les questionnaires par les participants en plus de faciliter l'exportation des données pour traitement analytique.
- Le CREDOC (3) a fait le choix d'un **hybridation** en laissant la possibilité aux participants de recourir soient aux carnets alimentaires « papier », soient à la saisie informatique afin de ne pas se couper d'une frange de la population moins « numérisée » (que ce soit pour des raisons générationnelles, de facilité d'utilisation du numérique, et/ou d'accès au numérique).
 - Ce choix permet ainsi une **meilleure représentativité** des populations participant à l'étude mais peut engendrer une hétérogénéité de la qualité des données collectées.
 - À l'instar des questionnaires de Nutrinet-Santé, **la saisie informatique guide le participant et le contraint dans le niveau de précision pour valider l'information** (par exemple au-delà des aliments du repas, le contexte est interrogé : seul, debout, devant la télé...)
 - **Ce niveau de précision n'est pas possible sur papier, tout comme il n'est pas possible contrôler « en temps réel » l'exhaustivité** de l'information sur le carnet papier. À cela peut venir s'ajouter la difficulté compréhension et de lecture des écritures sur papier.
 - Finalement, **les données remontées sur papier comportent le risque d'être de moindre qualité, plus hétérogènes et donc plus difficiles à analyser** par la suite (a minima, requérir un travail plus long de « mise au propre » des données).

La perte d'une compétence de la recherche : formation et accompagnement des participants

L'omniprésence et la démocratisation du numérique offre la possibilité de développer relativement simplement, rapidement et à grande échelle la participation d'une population à la phase d'observation d'un programme de recherche.

(2) S. Hercberg et al., « The Nutrinet-Santé Study: a web-based prospective study on the relationship between nutrition and health and determinants of dietary patterns and nutritional status », 2010

(3) Entretien avec une chargée d'études du CREDOC

« Aller chercher » les populations peu ou pas numérisées – à l'image de ce qui était la norme avant la numérisation de la société – **demande du temps, de l'argent mais aussi des compétences spécifiques** (1) :

- Il s'agit tout d'abord de trouver les bons relais sur le terrain et d'identifier les populations qui pourraient participer aux enquêtes
- Il faut ensuite les **convaincre, les motiver sur le temps de l'étude mais surtout, les former** pour que les informations qu'ils observent et renseignent correspondent au protocole mis en place :
 - Pour des enquêtes relatives à l'alimentation, il est nécessaire par exemple de prendre le temps d'expliquer les mesures des quantités, et comment les renseigner ensuite correctement dans les carnets d'alimentation papier

Ces compétences « terrain » ont tendance à disparaître : substituées d'une part par le recours (facilité) au numérique, et pâtissant d'autre part d'un moindre prestige et/ou intérêt en comparaison au travail d'analyse des données observées et collectées.

- **Cette perte de compétences est perçue comme un risque grandissant (2)**, rendant difficile voire mettant en péril sur le long-terme l'observation et la collecte de données au sein par exemple de populations peu ou pas numérisées.

(1) Entretien avec une chargée d'études du CREDOC

(2) Ibid

3 - LA NUMÉRISATION AU SERVICE DE LA RECHERCHE PARTICIPATIVE

3.1 - Capitaliser sur les possibilités du numérique pour faire avancer la recherche participative sur la biodiversité

L'omniprésence des technologies et outils numériques, doublée de leur large réappropriation pour tout à chacun, est perçue comme une opportunité pour les recherches participatives, particulièrement dans les domaines relatifs à la biodiversité.

Deux éléments de contexte viennent expliquer l'attrait pour les sciences participatives pour la recherche sur la biodiversité :

- La biodiversité est aujourd'hui devenue **un enjeu sociétal et de politique publique. Le besoin de connaissances sur la biodiversité s'en est vu accru** :
 - **des décideurs publics auprès des chercheurs** pour informer l'action publique
 - **des citoyens auprès des chercheurs**, demandeurs de mieux comprendre l'enjeu voire de participer à l'amélioration de sa compréhension
- **La numérisation de la société et la démocratisations de l'accès aux outils numériques démultiplient les possibilités d'observation volontaire et amateur** de la biodiversité. Ces pratiques sont historiquement bien ancrées avec notamment les communautés de naturalistes en France.

La conjonction des deux facteurs favorise ainsi le développement des programmes de sciences participatives (3)

- **La collecte de données est facilitée.** Les informations qui étaient avant consignées dans des carnets d'observation sont aujourd'hui possibles à être saisies en ligne.
- **Les données sont plus complètes et de meilleure qualité grâce aux outils numériques** (démocratisation de la photo numérique et plus généralement des appareils photographiques comme les pièges-photo).
- **Les échanges sont fluidifiés.** À l'instar du reste de la société, les échanges entre chercheurs et citoyens volontaires sont rapides, quasi instantanés et possibles via différents canaux.

(3) R. Julliard, « Science participative et suivi de la biodiversité : l'expérience Vigie-Nature », 2017

Les trois exemples suivants – The Wisconsin Bat Program, Vigie-Nature et The Penguin Watch-Zooniverse – nous permettent de comprendre le potentiel concret qu'offre le numérique aux programmes de recherche participative.

3.2 - Démocratisation des outils numériques : The Wisconsin Bat Program

La démocratisation des outils numériques a également permis de multiplier le type de données collectées :

- Les **saisies des données d'observation en ligne** sont venues remplacer les carnets d'observation papier
- **Sont venus s'y ajouter les « pièges-photo » et autres outils numériques** qui génèrent une quantité croissante de données hétérogènes à traiter.

Un premier exemple intéressant est celui du *Wisconsin Bat Program*. L'Etat du Wisconsin (USA) a une longue histoire de projets de sciences participatives incluant des citoyens dans les recherches menées sur la biodiversité. Il est notamment l'un des premiers Etats à participer au Christmas Bird Count en 1900 (1).

Depuis 2004 s'est formé le Wisconsin Citizen-based Monitoring Network, soutenu notamment par le Département des ressources naturelles de l'Etat du Wisconsin, qui réunit les différents programmes d'observation citoyenne.

Le programme dédié aux chauve-souris commence quant à lui en 2010 et les compétences de la communauté d'amateurs formés à l'observation des chauve-souris sont mobilisées notamment pour la recherche.

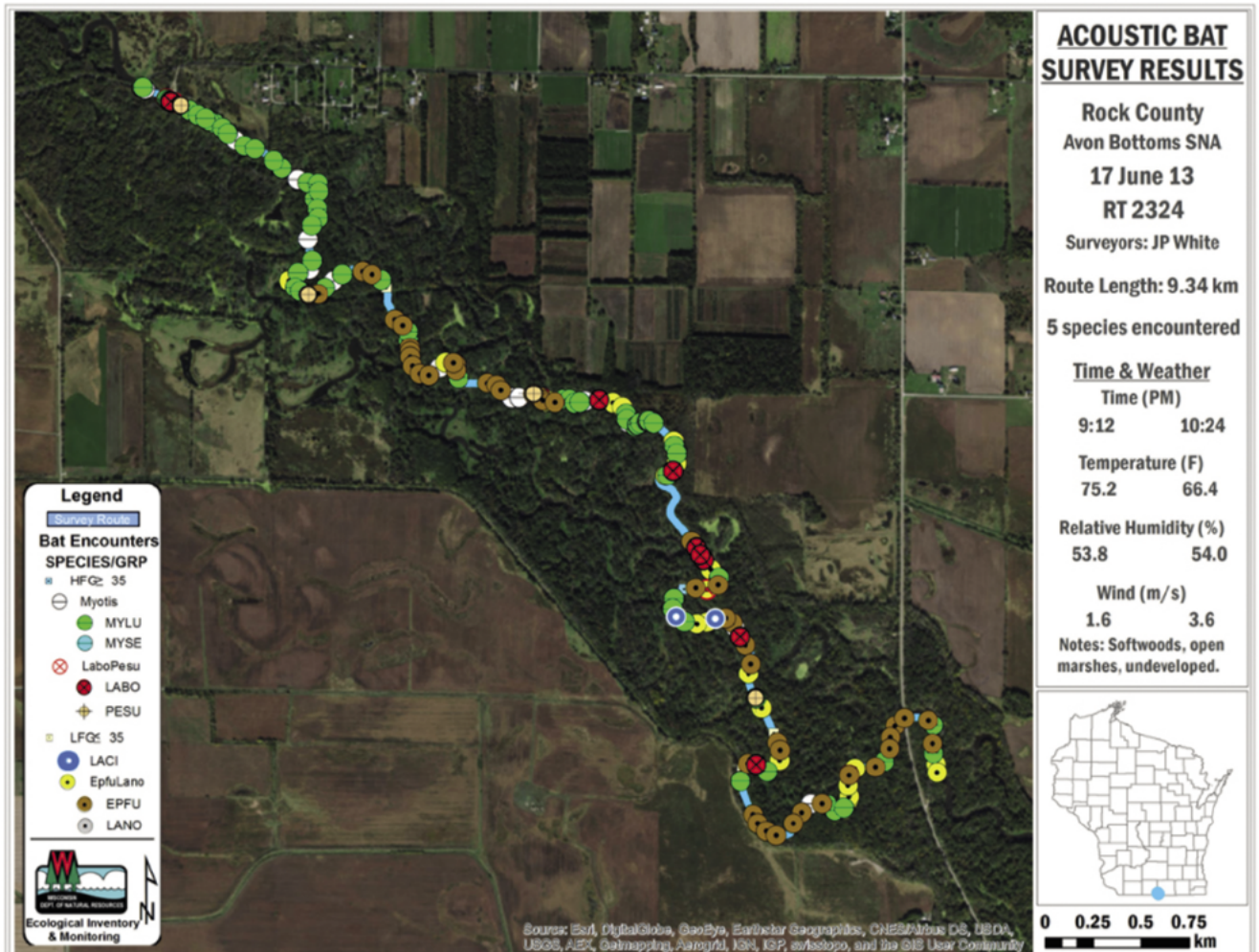
Des recherches (2) ont ainsi par exemple mobilisé la communauté d'amateurs afin qu'ils enregistrent via des capteurs ultra-sons connectés à la technologie GPS les activités des chauve-souris. Les données collectées permettent ainsi d'étudier leurs comportements et leur répartition géographique au sein de l'Etat.

(1) Projet de recherche participative d'observation citoyenne des oiseaux

EN SAVOIR PLUS

(2) A. F. Matzinger, « Bat activity in relation to landscape composition and spatial configuration in Wisconsin », 2011; The Wisconsin Bat Program,

EN SAVOIR PLUS



3.3 - Education des participations à la biodiversité : Vigie-Nature et l'OAB

Un deuxième exemple intéressant est celui de Vigie-Nature. Ce programme de **sciences participatives** (1) est porté par le Muséum national d'histoire naturelle (MNHN) en 1989 afin de rendre compte de l'évolution de la biodiversité et des pressions anthropiques dont elle fait l'objet.

Trois objectifs principaux sont fixés à Vigie-Nature : **1/** la production de données de terrain pour nourrir la recherche, **2/** l'accompagnement des politiques et **3/** la sensibilisation et la formation de la société civile.

Les nouvelles technologies numériques sont considérées comme facteur de succès de ce programme auprès des participants. La démocratisation par exemple des appareils photographiques et de dispositif comme le SPIPOLL permet aux participants de réaliser des « safari-photo » (2).

En France, 1000 participants ont ainsi produit sur 5 ans plus de 200 000 photos d'insectes et araignées.

Depuis 1989, les données et recherches menées par Vigie-Nature ont contribué à :

- Des publications scientifiques (en 2016 : 99 articles scientifiques, 12 thèses),
- Des rapports ministériels

La mise en place d'indicateurs de biodiversité nationaux (ONB) et européens (EBIGS; PECBMS).

(1) Vigie-Nature est caractérisé par un fonctionnement pluri-acteurs qui associe les sphères scientifiques et citoyennes, permis notamment par des relais associatifs locaux.
(2) R. Julliard, « Science participative et suivi de la biodiversité : l'expérience Vigie-Nature », 2017

2000 NAUTRALISTES ornithologues, entomologistes, botanistes...



STOC
Suivi temporel
des oiseaux
communs



SHOC
Suivi hivernal
des oiseaux
communs



Vigie-Chiro
Suivi des
Chauves-souris



STELI
Suivi temporel
des libellules



Vigie-Flore
Suivi des
plantes
communes



STERF
Suivi temporel
des rhopalocères
de France

14 000 PARTICIPANTS "GRAND PUBLIC"



**Observatoire de
la biodiversité
des jardins**



**Oiseaux des
jardins**



**Observatoire
des bourdons**



Spipoll
Suivi photogra-
phique des insectes
pollinisateurs



**Sauvage
de ma rue**



BioLit
Observatoire
du littoral

200 GESTIONNAIRES
D'ESPACES VERTS



6 000 ÉLÈVES



400 AGRICULTEURS



Au sein de Vigie-Nature, l'Observatoire agricole de la biodiversité (OAB) est destiné à observer la biodiversité en milieu agricole et selon les pratiques des parcelles observées (1).

L'objectif de l'OAB est double (2) :

Améliorer les connaissances scientifiques en alimentant une base de données et à terme en élaborant des indicateurs

Sensibiliser les agriculteurs participant à l'observation

À l'image de Vigie-Nature, l'Observatoire se caractérise par une dimension multi acteurs : les décisions sont par exemple prises conjointement entre partenaires scientifiques, techniques et institutionnels du monde agricole.

L'Observatoire suit 4 indicateurs significatifs (3) de la biodiversité et son évolution en milieu agricole en lien avec les pratiques culturales (voir infographie ci-contre).

Le protocole établi est un subtil équilibre entre exigence de la recherche et simplicité – cette dernière étant jugée cruciale dans l'appropriation des enjeux écologiques des agriculteurs pour remplir le second objectif (4)

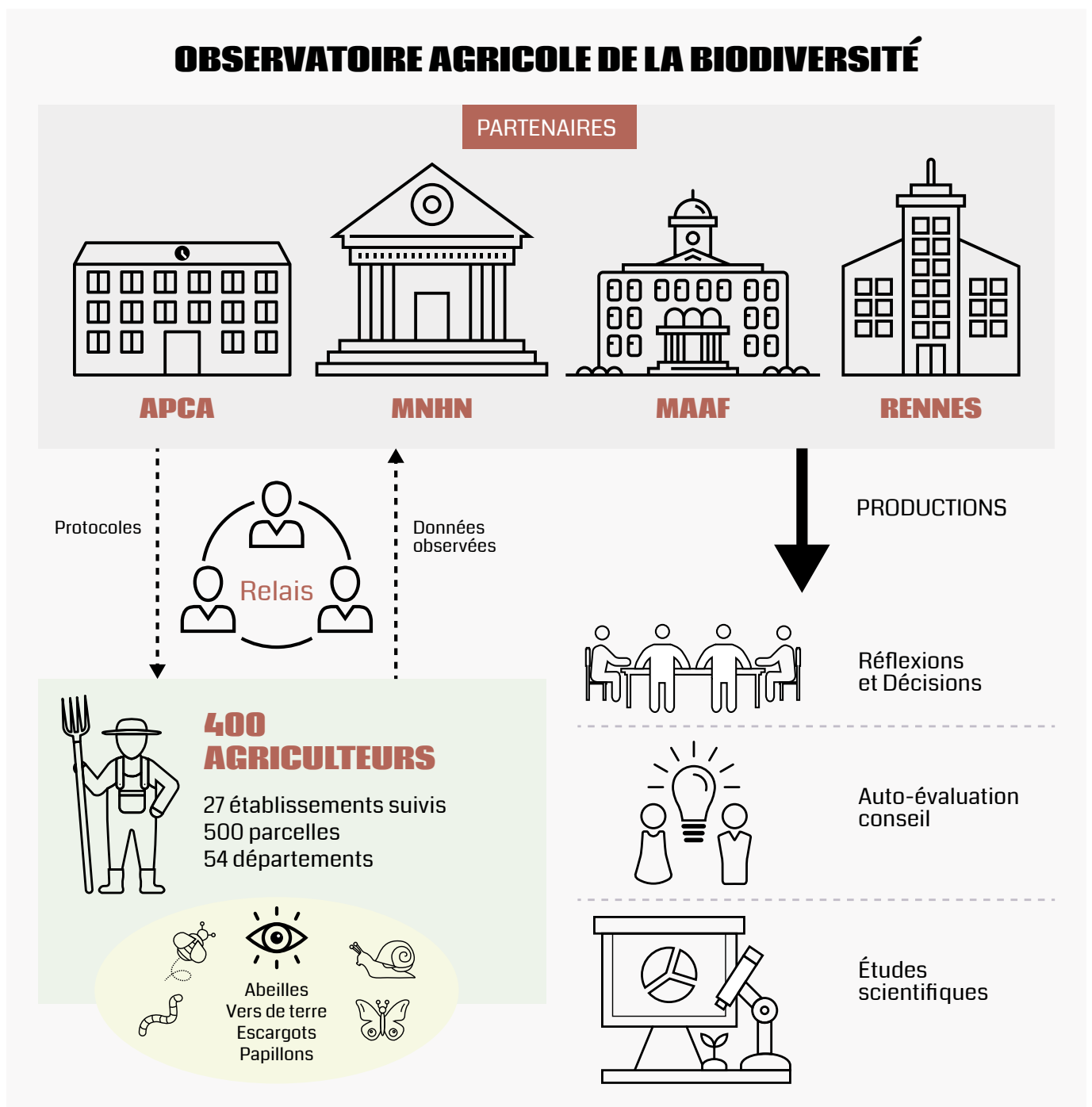
(1) observatoire-agricole-biodiversite.

EN SAVOIR PLUS

(2) R. Julliard, « Science participative et suivi de la biodiversité : l'expérience Vigie-Nature », 2017

(3) OAB, Bilan 2018, 2019

(4) R. Julliard, op. cit.



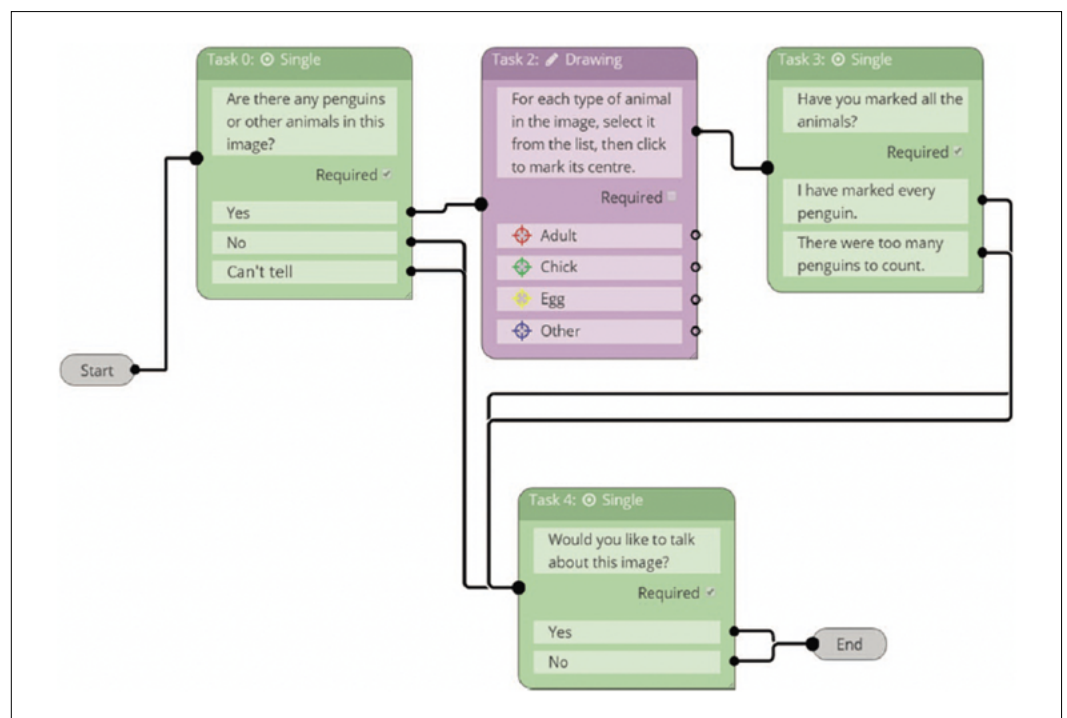
3.4 - Le traitement de la donnée par les participants : The Penguin Watch Project

Un troisième exemple est celui du programme The Penguin Watch de Zooniverse. Dans cet exemple, **les volontaires ne participent à la captation de la donnée mais sont au cœur de son traitement.**

Zooniverse est la plus grande et plus populaire plateforme mondiale de recherche participative (1): depuis son lancement en 2007, la plateforme vise à mettre en lumière des programmes de recherche qui seraient rendus impossibles (a minima, extrêmement complexes et coûteux) sans la participation citoyenne. Toutes les disciplines sont représentées sur la plateforme, donnant un large univers des possibles aux volontaires en fonction de leurs intérêts.

The Penguin Watch est un des projets proposés sur Zooniverse et vise à étudier les colonies de pingouins (notamment les îles Falkland et l'Antarctique). Leur observation est rendue possible par les nouvelles technologies numériques : différents pièges-photos connectés sont placés à des endroits stratégiques et prennent des séries de photos transmises aux chercheurs (2).

Ces derniers les mettent alors à disposition en ligne des volontaires selon une méthodologie dite du clustering (3) : les mêmes images sont soumises à différents volontaires et leurs identifications par clics des pingouins sur les images sont ensuite analysées.



Exemple d'interface pour un participant pour la reconnaissance d'animaux, The Penguin Watch

Exemple de reconnaissance croisée par différents participants sur une image, The Penguin Watch

(1) L.C. Johnson, « The Zooniverse », 2019

(2) F. M. Jones et al., « Data descriptor: Time-lapse imagery and volunteer classifications from the Zooniverse Penguin Watch Project », 2018

(3) Ibid

3.5 - Réflexions sur les conséquences

La qualité de la donnée

Une question récurrente qui se pose aux programmes de recherche participative est celle de l'**équilibre entre la surface de données collectables** (c'est-à-dire le nombre de personnes prêtes à participer) et la **qualité des données non-expertes collectées** (1) :

- Aux pôles opposés, il y a par exemple :
 - **La recherche participative plus « opportuniste »** : l'observation de l'information est peu contraignante afin d'en faciliter la saisie et la remonter. C'est par exemple le cas de l'application Plantnet rattachée au programme de recherche FlorisTic (2)
 - **La recherche participative structurée par des protocoles stricts et contraignants**, à l'image des protocoles de recherche « classique ». C'est par exemple le cas des protocoles des différents observatoires de Vigie-Nature (voir ci-après l'exemple du protocole de l'Observatoire agricole de la biodiversité)
- Les termes de l'équilibre peuvent ainsi être synthétisés comme suit :
 - Un **protocole d'observation peu contraignant** peut permettre de récolter plus de données mais dont la **qualité pourrait être moins bonne et demandée plus de traitement et vérification ultérieurs**
 - Un protocole d'observation contraignant peut quant à lui assurer une qualité de donnée élevée mais rebuter certains participants, voire les démotiver sur le temps long

(1) Entretien avec une chercheuse en biodiversité et sciences participatives

(2) Présentation du programme de recherche FlorisTic

EN SAVOIR PLUS

Portée pédagogique et éducative de la participation

Les programmes de recherche participative ont le double objectif de **collecte de données et de sensibilisation voire d'éducation des participants** aux domaines de recherche investigués :

- Cette deuxième dimension est perçue comme étant un enjeu important des différents exemples de programmes de recherche de biodiversité décrits précédemment
- **Si les participants de ces programmes de recherche participative sur la biodiversité sont souvent des acteurs déjà sensibilisés à ces questions** (au premier titre desquels les communautés de naturalistes), **la participation peut permettre d'aller un cran plus loin** et d'avoir un impact concret par exemple sur les pratiques individuelles de jardinage (3)

L'Observatoire agricole de la biodiversité est un cas un peu à part puisqu'il cible un public particulier : en l'occurrence, les **agriculteurs**. Leur participation à l'Observatoire est jugée importante sur différents niveaux (4) :

- **Valoriser les agriculteurs** dans leur rôle de « **sentinelles** » de la biodiversité,
- Leur permettre d'**acquérir de nouvelles connaissances sur la biodiversité** via la participation à l'observatoire,
- **Relier les observations réalisées sur la biodiversité sur leurs parcelles à leurs pratiques agricoles.**

Participer à l'Observatoire permet ainsi aux agriculteurs d'être acteur de l'observation de la biodiversité mais aussi de la **réflexion sur leurs pratiques** (5).

- Cette réflexion est d'autant plus intéressante qu'elle est partagée entre les acteurs participants et objectivées par les données relevées au fil des années sur les parcelles.
- La participation à cet observatoire des agriculteurs – et plus largement les programmes de recherche participative – ont un potentiel transformatif en contribuant à l'« **encapacitation** » des **acteurs** qui comprennent mieux leur environnement et peuvent adapter leurs actions en conséquence (6).

(3) Entretien avec une chercheuse en biodiversité et sciences participatives

(4) R. Julliard, op. cit.

(5) O. Billaud, « Du tracteur au carnet de comptage, un projet de sciences participatives 100 % agricole », 2020

(6) R. Julliard, op. cit.

Le traitement de la donnée : l'humain, le machine *learning* et l'intelligence artificielle

La masse et l'hétérogénéité des données collectées rendent caduc le « simple » traitement par l'être humain et oblige au développement de technologies de stockages et d'outils de traitement :

- Des **collaborations de recherche** se créent entre chercheurs spécialisés dans un domaine avec des chercheurs en développement numérique et informatique :
 - C'est l'exemple du **programme ANR D2KAB** (1) dont la question de départ est celle de la **capacité des chercheurs à traiter et analyser seuls les volumes croissants de données produites de plus en plus rapidement en agronomie et biodiversité**. Le programme de recherche ambitionne ainsi de **conjuguer les compétences** de différents laboratoires en informatique, en écologie, en écosystèmes etc. afin de **développer des processus permettant de traiter les données pour les transformer en connaissances manipulables** par des chercheurs de différentes disciplines.
- Plus particulièrement, **des technologies numériques comme celles du machine learning (ML) et de l'intelligence artificielle (AI) sont de plus en plus utilisées dans la recherche** :
 - Les développements récents des « **deep neural networks** » (réseaux neuronaux profonds) ont rendu ces technologies extrêmement performantes : plutôt que de calculer toutes les possibilités et d'utiliser les statistiques pour déterminer la meilleure, cette technologie permet aux machines de créer leur propre intuition et de comprendre quelles possibilités sont meilleures que d'autres dans ce contexte.
 - **Leur performance dépasserait aujourd'hui celles des êtres humains** : en témoigne par exemple la découverte en 2017 de deux exoplanètes par DeepMind Artificial Intelligence développée par Google et utilisée en partenariat avec la NASA (2) qui n'avaient pas été identifiées par l'œil humain.

La question posée est alors celle de **l'articulation entre les compétences humaines et celles des technologies numériques**, d'autant plus à enjeux dans le cadre des programmes de recherches participatives.

Pour l'instant, **les différents programmes par exemple de Zooniverse ont décidé d'une hybridation permettant de tirer le meilleur parti des deux** :

D'une part, **recourir aux technologies numériques du ML et de l'AI afin de ne pas solliciter les participants là où les technologies seront plus efficaces qu'eux** (3),

D'autre part, **continuer à bénéficier et capitaliser sur l'intelligence humaine pour « entraîner » les machines**. En effet, ces « réseaux neuronaux profonds » requièrent de grands ensembles de données (images, sons et autres) afin de s'entraîner. Ces grands ensembles de données doivent avoir été au préalable soigneusement et précisément identifiées et catégorisées. **Contribuer à l'apprentissage des machines en participant à fournir des ensembles de données de haute qualité avant qu'elles ne soient « autonomes » est ainsi perçu comme un nouvel axe important des recherches participatives** (4).

Néanmoins, il reste un pan de la reconnaissance (que ce soit d'exoplanète ou d'espèce) pour lequel les technologies numériques et les machines ne peuvent remplacer (encore) l'être humain : **la détection des anomalies**. Aussi puissantes soient-elles, les machines n'arrivent pas (encore) à date à substituer l'œil et l'intelligence de l'être humain dans sa capacité à **identifier des éléments inhabituels et inattendus** :

C'est alors une nouvelle **piste de complémentarité entre les intelligences humaine et artificielle** : là où la machine atteint ses limites, l'être humain – et par exemple les participants aux programmes de recherche participative – aide à identifier la cause de la « confusion » de la machine et à y trouver une réponse (5).

(1) « Des données aux connaissances en agronomie et biodiversité – D2KAB » :

EN SAVOIR PLUS

(2) J. Clark, « Google's artificial intelligence finds two new exoplanets missed by human eyes », 2017

(3) UW, « Researchers use artificial intelligence to identify, count, describe wild animals », 2018

(4) K. Schawinski, « Beyond today's crowdsourced science to tomorrow's citizen science cyborgs », 2016

(5) Ibid

L'hybridation est perçue finalement comme la voie à poursuivre : **si les programmes de recherche nécessitant un traitement massif des données n'avaient plus du tout recours à l'être humain mais uniquement à la machine, ils hypothéqueraient alors la dimension de sensibilisation, d'éducation et de pédagogie jugée tout aussi importante que celle de la connaissance** – notamment sur la question de la préservation de la biodiversité (1)

4 - NUMÉRISATION DE LA RECHERCHE : ABSENCE ET SURABONDANCE DE DONNÉES

4.1 - Dépendance de la recherche à l'existence des données et à leur traitement par les outils numériques

Malgré l'omniprésence des technologies et outils numériques et leur large réappropriation dans le grand public, **tous les domaines de recherche potentielle ne sont pas couverts uniformément**. Autrement dit, il existe de fortes disparités dans la génération de données rendues disponibles via les outils numériques en fonction des sujets.

Un domaine par exemple aujourd'hui peu et mal couvert en France est celui de la remontée d'informations sur les cancers, en particulier ceux pédiatriques, et les maladies chroniques.

- En France métropolitaine, 14 registres des cancers couvrant 19 départements existent, soit une couverture de 24% de la population métropolitaine (2)
- Les quelques données complémentaires pouvant exister émanent des services hospitaliers eux-mêmes lorsque ces derniers décident d'investiguer les causes des diagnostics qu'ils posent (3)

Alors que les données consolidées et vérifiées, notamment par les autorités publiques, peuvent manquer, **d'autres domaines au contraire font l'objet d'une (trop) grande abondance de travaux de recherche** – publique, privée voire également citoyenne.

- **L'impact des produits phytosanitaires sur les milieux** est un champ d'études vaste et complexe aujourd'hui investi par une multitude d'acteurs venant des sphères académiques publiques, des secteurs économiques privés (notamment ceux directement liés à la fabrication des produits phytosanitaires) et des citoyens (4).
- **Ce champ d'études est d'autant plus investi qu'il a trait à des enjeux économiques, politiques et de santé publique (entre autres)**.

Les exemples développés ci-après permettent d'investiguer les réponses apportées pour tenter de compenser l'absence comme la surabondance de données.

4.2 - Pallier le manque de données : les collectifs citoyens sur les produits phytosanitaires

Pour pallier ce manque de données d'ordre de santé publique, notamment sur les questions liant les développements de cancers pédiatriques et les utilisations de produits phytosanitaires, des **collectifs citoyens** s'organisent :

- Un des axes de travail de ces collectifs est de **documenter, recenser et étayer** les « intuitions » de causes à effets entre l'exposition à des produits phytosanitaires et des développements de maladies, dont les cancers. Ce travail s'appuie (5) :
 - D'un côté, sur **une collecte de données via les réseaux** : de l'observation quantifiée et consolidée,

(1) Entretien avec une chercheuse en biodiversité et sciences participatives

(2) Liste des départements métropolitains couverts, auxquels il faut ajouter 3 registres ultramarins :

EN SAVOIR PLUS

(3) Entretien avec une chercheuse en toxicologie ; à ce sujet, lire I. Léraud et P. Van Hove, « Algues vertes, l'histoire interdite », 2019

(4) Ibid

(5) Entretien avec une chercheuse en toxicologie

- De l'autre, sur une **analyse comparative** entre les données observées et consolidées via leurs réseaux et des données comparables – si celles-ci existent et sont accessibles,
- L'objectif est d'étayer **la demande de reconnaissances des cas en étudiant leur prévalence statistique**.

Ces observations et recherches menées par les collectifs citoyens ont parfois rencontré un « succès » et ont permis **la mise en place d'études scientifiques par les autorités publiques** (1), même si les périmètres des études ne donnent souvent pas entière satisfaction (cas retenus, période de temps etc.) (2).

- Un exemple unique en son genre est celui de **l'Institut écocitoyen** (3) :
 - Fondé en 2010 sous l'impulsion de la mairie de Fos-sur-Mer et de la communauté d'agglomération Ouest-Provence suite à l'observation de développements de cancers dans une zone très industrielle
 - Travaux de l'Institut s'adressent aux acteurs du territoire en vue de la réduction des émissions polluantes, de l'adaptation des suivis environnementaux aux polluants spécifiques, et de la réhabilitation des sites contaminés.
 - Résultats utilisés par le Comité d'évaluation et de contrôle des politiques publiques de l'Assemblée nationale, ainsi que par le Conseil Général de l'Environnement et du Développement Durable.

Au global, il est considéré que les travaux d'observation et de documentation réalisés par ces collectifs citoyens **sont une source d'information qui pourrait (devrait) être mobilisée de façon plus systémique par les acteurs de la recherche là où les données manquent ou sont à compléter** (4) :

- À titre d'exemple, il serait possible de recouper des informations d'observations citoyennes et des carnets d'épandage pour contribuer à documenter les risques des expositions aux produits phytosanitaires

4.3 – Aider à l'interprétation des données : les Adverse Outcome Pathways

À l'image du recours au machine learning et à l'intelligence artificielle dans le traitement des données collectées au sein des programmes de recherche participative en biodiversité, il existe d'autres **procédés s'appuyant sur les possibilités des technologies numériques pour traiter et analyser les données sur les impacts des produits phytosanitaires sur les milieux**.

À titre d'exemple, face à la génération d'une vaste quantité de données (éco)toxicologiques, **l'Adverse Outcome Pathway (AOP)** a été développé pour organiser les données d'une manière accessible et profitable aux évaluateurs des risques (5) :

- Construction conceptuelle visualisant les connaissances disponibles sur les liens entre un événement déclencheur au niveau moléculaire et un effet néfaste sur l'organisme (6)
- Principal atout de l'AOP est d'évaluer l'impact d'une substance chimique sur l'organisme et la population
 - **Capitaliser sur le croisement des connaissances déjà produites dans le cadre de recherches menées par différentes disciplines et à travers le monde (Data mining)**
 - **Réduire et cibler le recours aux expérimentations sur les animaux de laboratoire (7)**

Afin de prioriser la prise de décision au regard du très grand nombre de produits chimiques présents sur le marché, **l'AOP est adopté au-delà des sphères scientifiques : par les autorités publiques, internationales, et/ou autorisant la mise en marché des produits** (8)

(1) Assemblée nationale, Commission d'enquête sur l'évaluation des politiques publiques de santé environnementale, 2020

(2) Collectif SCE, « Cluster et études épidémiologiques en France » ; Collectif SCE, « Le cluster de cancers pédiatriques sur le territoire autour de Sainte Pazanne », 2020

(3) Reporterre, « Fos-sur-Mer : un projet unique de science participative pour mesurer les pollutions industrielles », 2020 ; Pour en savoir plus sur l'Institut écocitoyen :

EN SAVOIR PLUS

(4) Entretien avec une chercheuse en toxicologie

(5) AINERIS, Méthodes et outils émergents d'évaluation des dangers », 2016

(6) ECOTOX-INRAE, « Adverse Outcome Pathway : concept et exemple », 2016

(7) PAN, « AOP: The Trojan Horse for Industry Lobby Tools », 2016

(8) T. Dumas, « Les approches - omiques, métabolomique et protéomique, pour l'étude de la relation de cause à effet entre contaminants émergents, produits pharmaceutiques et organismes marins », 2020

(1) Entretien avec
une chercheuse en
écotoxicologie

(2) « Des données aux
connaissances en
agronomie et biodiversité
- D2KAB » :

EN SAVOIR PLUS

4.4 - Réflexions sur les connaissances

La limitation des connaissances par le filtre du numérique

Bien qu'à deux extrêmes opposés, la (sur)abondance ou la (quasi) absence de données posent la même question : celle du filtre posé par la numérisation.

- **La donnée qui n'est pas numérisée est une donnée qui « n'existe pas » ou moins que celles numérisées.** Il y a un enjeu alors à donner de la visibilité à des sujets et objets de recherche en créant de la donnée dessus, notamment via les outils numériques
 - La numérisation de la remontée d'informations éparpillées au sein des territoires et des services hospitaliers sur les cancers notamment pédiatriques, croisée à de la donnée numérisée sur les pratiques agricoles (notamment relatives à l'utilisation de produits phytosanitaires) serait perçue comme une révolution pour les recherches en écotoxicologie (1)
- **À l'inverse, les données produites en grande quantité par les outils numériques acquièrent une grande visibilité, entraînant un cercle de massification rendant difficile le passage de « donnée » à « connaissance » (2).**
 - L'enjeu devient alors de développer des outils qui permettent de traiter la donnée afin de la rendre compréhensible et permettre son analyse pour qu'elle se transforme en connaissance
 - Mais l'outil, par exemple l'AOP, s'il peut aider à analyser peut aussi entraîner des **biais** :
 - **L'AOP ne fait pas remonter l'information sur la source de la donnée.** Ce manque de regard sur la source est considéré comme étant problématique dans le contexte de la recherche sur les impacts des produits phytosanitaires parasitée par **des « recherches leurres »** (c'est à dire des recherches tout ou partiellement réalisées par des acteurs en situation de conflit d'intérêt avec les fabricants de produits phytosanitaires)
 - **L'AOP offre un traitement quantitatif « neutre » qui ne permet pas d'avoir un regard qualitatif sur la source de la donnée**
 - **Les connaissances produites via le traitement AOP peuvent ainsi être parasitées par données générées par des recherches leurres. Les connaissances s'en trouvent limiter.**

BIBLIOGRAPHIE

Des données aux connaissances en agronomie et biodiversité – D2KAB :
<https://d2kab.mystrikingly.com/>

Assemblée nationale, *Commission d'enquête sur l'évaluation des politiques publiques de santé environnementale*, 2020

Billaud, *Du tracteur au carnet de comptage, un projet de sciences participatives 100 % agricole*, 2020

Chantal J. et al., *Contribution of ultra-processed foods in the diet of adults from the French NutriNet-Santé study*, 2017

Clark, *Google's artificial intelligence finds two new exoplanets missed by human eyes*, 2017

CNAM, *Les nouveaux résultats de l'enquête Nutrinet-Santé décryptent plus précisément les comportements alimentaires des français*, 2011

Cochoy et al., *Digitalizing consumer society: equipment and devices of digital consumption*, 2020

Collectif SCE, *Le cluster de cancers pédiatriques sur le territoire autour de Sainte Pazanne*, 2020

CREDOC, *Comportements et attitudes alimentaires en France – CAF 2021*, 2021

Dumas, *Les approches – omiques, métabolomique et protéomique, pour l'étude de la relation de cause à effet entre contaminants émergents, produits pharmaceutiques et organismes marins*, 2020

ECOTOX-INRAE, *Adverse Outcome Pathway : concept et exemple*, 2016

Fardet et E. Rock, *Ultra-processed foods and food system sustainability: what are the links?*, 2020

Floris'Tic : <http://floristic.org/>

Hercberg et al., *The Nutrinet-Santé Study: a web-based prospective study on the relationship between nutrition and health and déterminants of dietary patterns and nutritional status*, 2010

INERIS, *Méthodes et outils émergents d'évaluation des dangers*, 2016

Institut écocitoyen : institut-ecocitoyen.fr/pres.php

Johnson, *The Zooniverse*, 2019

Jones et al., *Data descriptor: Time-lapse imagery and volunteer classifications from the Zooniverse Penguin Watch Project*, 2018

Julliard, *Science participative et suivi de la biodiversité : l'expérience Vigie-Nature*, 2017

Léraud et P. Van Hove, *Algues vertes, l'histoire interdite*, 2019

Loupiac et C. Blutz, *Au croisement du marketing et de la comptabilité : l'essor des activités d'évaluation et de contrôle par le consommateur*, 2020

Matzinger, *Bat activity in relation to landscape composition and spatial configuration in Wisconsin*, 2011

Ngueilbaye et al., *Adoption of human metabolic processes as Data Quality Based Models*, 2020

OAB, *Bilan 2018*, 2019

PAN, *AOP: The Trojan Horse for Industry Lobby Tools*, 2016

Ramirez-Portilla et al., *The era of Open Food? Exploring the influence of open and collaborative innovation in the food industry*, 2016

Reporterre, *Fos-sur-Mer : un projet unique de science participative pour mesurer les pollutions industrielles*, 2020

Réseau PROSPER, *Big & Open data en recherche à l'horizon 2040*, 2019

Schawinski, *Beyond today's crowdsourced science to tomorrow's citizen science cyborgs*, 2016

Soutjis, *Gouverner la qualité alimentaire par les applications*, 2020

The Wisconsin Bat Program, <https://wiatri.net/inventory/bats/>

UW, *Researchers use artificial intelligence to identify, count, describe wild animals*, 2018